ELSEVIER

Linear Algebra and its Applications 305 (2000) 173–185

LINEAR ALGEBRA AND ITS APPLICATIONS

www.elsevier.com/locate/laa

Difference equations in a general setting

D.J. Hartfiel

Department of Mathematics, Texas A&M University, College Station, TX 77843, USA

Received 30 June 1998; accepted 10 October 1999

Submitted by M. Neumann

Abstract

For a square matrix A, the difference equation $y_{k+1} = y_k A$ generates a sequence of vectors $y_0, y_1, \ldots$ The long run behavior of this sequence is described in terms of a group obtained from $A, A^2 \ldots$ The results are given in a topological semigroup setting so they can also be applied to more general difference equations. © 2000 Elsevier Science Inc. All rights reserved.

AMS classification: 15A24

Keywords: Difference equations; Solution behavior; Topological semigroup

Let A be an $n \times n$ matrix. The classical difference equation

$$y_{k+1} = y_k A,$$

where y_0 is a vector, arises in numerous applied problems. (See [6].)

The solution to this difference equation need not converge. In those cases, and for other reasons, it is sometimes useful to consider a difference equation

$$S_{k+1} = S_k A,$$

where S_0 is a set of vectors and $S_{k+1} = \{xA\colon\ x \in S_k\}$. For example, if

$$S_0 = \operatorname{convex}\{(1,0,0), (0,1,0), (0,0,1)\},$$

the set of all probability distributions, it can be shown that for

$$A = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

E-mail address: hartfiel@math.tamu.edu (D.J. Hartfiel).

0024-3795/00/$ - see front matter
PII: S 0 0 2 4 - 3 7 9 5 (9 9) 0 0 2 3 8 - 4

$S_0, S_1, \ldots$ converges, in the Hausdorff metric, to convex$\{(1,0,0), (0,\frac{1}{2},\frac{1}{2})\}$. This occurs even though the classical difference equation $y_{k+1} = y_k A$ need not converge for particular probability distributions. (See [2,3] for more details.)

An even more general way to study difference equations is to consider

$$S_{k+1} = S_k M,$$

where S_0 is a set of vectors, M is a set of $n \times n$ matrices, and $S_{k+1} = \{xA\colon\ x \in S_k$ and $A \in M\}$. Barnsley [1] uses this approach to construct fractals and Hartfiel [4] uses it to study Markov chains with imprecise data.

What we intend to show in this paper is that the behavior of solutions to difference equations, in these various settings, can be described using a group obtained from the sequence $A, A^2, \ldots$ or $M, M^2, \ldots$

To do this, let p be an $n \times n$ real matrix or a set of such matrices. Let W be a subset of $\mathbb{R}^n$. We consider

$$W, Wp, Wp^2, \ldots$$

By using topological semigroups, we intend to describe the long run behavior of the sequence.

Results. Let S be a compact topological semigroup (S is a Hausdorff space and multiplication in S is continuous) whose topology is determined by a continuous metric d. Some examples of such follow.

(i) The semigroups of the $n \times n$ stochastic matrices and the $n \times n$ substochastic matrices: Here we use the 1-norm for vectors and the induced matrix norm $\|\cdot\|_1$, under left multiplication by those vectors,

$$\|A\|_1 = \max_{x \neq 0} \frac{\|xA\|_1}{\|x\|_1}.$$

Recall that

$$\|A\|_1 = \max_i \sum_{k=1}^{n} |a_{ik}|$$

and

$$d(A, B) = \|A - B\|_1.$$

(ii) The semigroup of $n \times n$ orthogonal matrices: Using the 2-norm for vectors and the induced matrix norm $\|\cdot\|_2$ we have

$$d(A, B) = \|A - B\|_2.$$

(iii) Special 0-pattern matrices: Let $\mathbb{R}^{n \times n}$ be the set of $n \times n$ $(n > 1)$ real matrices with the induced norm $\|\cdot\|_\infty$ for left multiplication by vectors. Thus,

$$\|A\|_\infty = \max_j \sum_{k=1}^{n} |a_{kj}|$$

and

$$d(A, B) = \|A - B\|_\infty.$$

Let S be the subset of $\mathbb{R}^{n\times n}$ defined by

$$S = \left\{A\colon\ A = \begin{bmatrix} B & 0 \\ b & 1 \end{bmatrix} \text{ and } \|A\|_\infty \leqslant 1\right\}.$$

Then S is a compact topological semigroup.

(iv) Finite generating sets: There are numerous variants of the semigroup given in (iii). We look at one of these.
Barnsley uses the affine maps

$$x_{k+1} = x_k \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} + (1,1); \quad x_{k+1} = x_k \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} + (1,50);$$

$$x_{k+1} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} x_k + (50,50)$$

to construct the Sierpinski triangle. (Other fractals require other affine maps.) For a given triangle T_0, Barnsley computes

$$T_{k+1} = \left(T_k \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} + (1,1)\right) \cup \left(T_k \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} + (1,50)\right)$$
$$\cup \left(T_k \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} + (50,50)\right)$$

$$\left(T_k \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} + (1,1) = \left\{x \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} + (1,1)\colon\ x \in T_k\right\}, \text{etc.}\right)$$

and shows $T_0 \cup T_1, \cup \cdots$ gives the Sierpinski triangle.
To put this construction into our difference equation format set

$$A_1 = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 1 & 50 & 1 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 50 & 50 & 1 \end{bmatrix}$$

and $y_k = (x_k, 1)$. Thus, using these homogeneous coordinates, we have

$$S_{k+1} = S_k A_1 \cup S_k A_2 \cup S_k A_3 = S_k\{A_1, A_2, A_3\},$$

where $S_k = \{(x,1)\colon\ x \in T_k\}$. In this notation

$$S_0 \cup S_1 \cup \cdots$$

gives the Sierpinski triangle.

To obtain a semigroup, let $N = \{I, A_1, A_2, A_3\}$. Define

$$N^2 = \{C_{i_1}C_{i_2}\colon\ C_{i_1}, C_{i_2} \in N\},$$
$$N^3 = \{C_{i_1}C_{i_2}C_{i_3}\colon\ C_{i_1}, C_{i_2}, C_{i_3} \in N\},$$
$$\ldots$$

Then

$$S = N \cup N^2 \cup N^3 \cup \cdots$$

is a semigroup.

For the norm we have

$$\|A\| = \|DAD^{-1}\|_\infty,$$

where

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.001 \end{bmatrix},$$

to assure $\|C\| \leqslant 1$ for all $C \in N$ and thus for all $C \in S$. Using this norm on $\mathbb{R}^{3\times 3}$, $\overline{S}$, the closure of S, is a compact semigroup.

Finally, if

$$C_k = S_0 \cup \cdots \cup S_k \quad \text{then}$$
$$C_{k+1} = C_k N$$

and C_k approaches the Sierpinski triangle. This puts fractal construction of Barnsley in our setting.

(v) A semigroup of compact sets: Given a norm $\|\cdot\|$ on the set $\mathbb{R}^{n\times n}$ of $n \times n$ matrices and that $d(A, B) = \|A - B\|$, let S be any compact semigroup of $\mathbb{R}^{n\times n}$. Define

$$H(S) = \{X\colon\ X \text{ is a nonempty compact set in } S\}.$$

Using the product of sets, $H(S)$ is a semigroup. Further, with the Hausdorff metric, $H(S)$ is a compact semigroup (see Barnsley [1]).

The semigroup $H(S)$, where S is the set of stochastic matrices, is used in Hartfiel [4], a study concerning imprecise transition matrices, for Markov chains.

We now develop the general results of the paper. Notationally, we let S be a compact topological semigroup with continuous metric d and $p \in S$. Consider the sequence $p, p^2, p^3, \ldots$. It is shown in [5] that the set G of limit points of this sequence is a group.

Before giving the major work of this paper, we prove two results about G.

Lemma 1. *Define $\pi : G \to G$ by $\pi(g) = gp$. Then π is an isomorphism.*

Proof. To show that $\pi : G \to G$, let $g \in G$. Then there is a sequence $p^{i_1}, p^{i_2}, \ldots$ which converges to g. Since multiplication is continuous in a topological semigroup, $p^{i_1+1}, p^{i_2+1}, \dots$ converges to gp. Thus, gp is a limit point of $p, p^2, \dots$ and so $gp \in G$.

To show π is onto, let $g \in G$. Then, g is the limit of some sequence $p^{i_1}, p^{i_2}, \dots$ Consider $p^{i_1-1}, p^{i_2-1}, \dots$ and take a convergent subsequence of it say $p^{j_1-1}, p^{j_2-1}, \dots$ Suppose this sequence converges to $\bar{g} \in G$. Then $\bar{g}p$ is the limit point of $p^{j_1}, p^{j_2}, \dots$ which is g. Thus, $\bar{g}p = g$ and $\pi\colon\ G \to G$ is onto.

Finally, to show that π is one-to-one, suppose that $g, \bar{g} \in G$ and that $gp = \bar{g}p$. Let e be the identity element of G. Then $g(ep) = \bar{g}(ep)$. Since $\pi\colon\ G \to G$, $ep \in G$ and thus has an inverse there. Hence, $g = \bar{g}$ which shows that p is one-to-one. $\square$

Define the coefficient of contraction (or contractive factor) for p as

$$\mathscr{T}_p = \sup_{x,y} \frac{d(xp, yp)}{d(x, y)},$$

where the sup is over all distinct $x, y \in \mathbb{R}^n$. We are interested in those semigroups in which $\mathscr{T}_p \leqslant 1$. For the stochastic matrix semigroup, the substochastic matrix semigroup, the orthogonal matrix semigroup, and the semigroups of (iii) and (iv), $\mathscr{T}_p \leqslant \|p\|$ and $\|p\| \leqslant 1$. Further, for $H(S)$ defined in (v), $\mathscr{T}_p \leqslant 1$ is shown in [4].

Using this notation we give a condition that assures G is finite. Later we will show that G determines the behavior of difference equations, as described in the introduction, and when G is finite that behavior is more easily calculated.

Lemma 2. *Suppose $\mathscr{T}_p \leqslant 1$. Then G is finite if and only if there is a positive integer s and an element $g \in G$ such that $gp^s = g$.*

Proof. By Lemma 1, the direct implication is clear. Thus, we need only argue the converse implication. For this, we let g denote an element as described in the converse.

Set $g_0 = g$, $g_1 = gp, \dots, g_{s-1} = gp^{s-1}$, $gp^s = g$, where s is the smallest such integer. We show $G = \{g_0, \dots, g_{s-1}\}$. The proof is by contradiction.

Suppose there is a $g_s \in G$ and $g_s \neq g_k$ for $k = 0, \dots, s-1$. Let $\epsilon = \frac{1}{3} \min d(g_i, g_j)$. (The minimum is taken over i, j where $0 \leqslant i \neq j \leqslant s$.) Let $p^{i_1}, p^{i_2}, \dots$ be a subsequence of $p^1, p^2, \dots$ which converges to g. Let $K > 0$ such that if $k > K$ then $d(p^{i_k}, g) < \epsilon$. Then if t is a positive integer and $t = sq + r$ where $0 \leqslant r < s$, $d(p^{i_k+t}, g_r) = d(p^{i_k+t}, gp^r) = d(p^{i_k}p^{sq+r}, gp^{sq+r}) \leqslant d(p^{i_k}, g) < \epsilon$. Thus, for $i > i_K$, $d(p^i, g_s) > \epsilon$ and so g_s is not a limit point of $p^1, p^2, \dots$. Hence $g_s \notin G$, a contradiction from which the result follows. $\square$

For the major work we let W be a compact subset of $\mathbb{R}^n$ and p an $n \times n$ matrix (in a compact topological semigroup S in $\mathbb{R}^{n\times n}$) or a compact set of $n \times n$ matrices (in a compact topological semigroup as described in (v)).

Consider the sequence of compact sets $Wp, Wp^2, \dots$ Define the limit set W_∞ as

$$W_\infty = \{s \in W\colon\ s \text{ is a limit point of } w_1 p, w_2 p^2, \dots \text{ where } w_1, w_2, \dots \text{ are in } W\}. \tag{1}$$

It is easily seen that W_∞ is compact. We now show that it is G that determines W_∞ from W. This description is below.

Theorem 1. $W_\infty = \bigcup_{g\in G} Wg$.

Proof. Let $u \in W_\infty$. Then there is a sequence $w_{i_1}p^{i_1}, w_{i_2}p^{i_2}, \dots$ which converges to u. Let $w_{j_1}, w_{j_2}, \dots$ be a subsequence of $w_{i_1}, w_{i_2}, \dots$ that converges, say to w. Since W is compact, $w \in W$. Let $p^{r_1}, p^{r_2}, \dots$ be a subsequence of $p^{j_1}, p^{j_2}, \dots$ which converges to say g. Thus, $g \in G$. Now, $u = \lim_{k\to\infty} w_{r_k}p^{r_k} = wg \in Wg$. Hence $W_\infty \subseteq \bigcup_{g\in G} Wg$.

Conversely, let $u \in \bigcup_{g\in G} Wg$. Then, $u \in Wg$ for some $g \in G$. Thus, $u = wg$ for some $w \in W$. Since $g \in G$, there is a sequence $p^{i_1}, p^{i_2}, \dots$ which converges to g. Thus, $wp^{i_1}, wp^{i_2}, \dots$ converges to wg. And hence, since $wg = u$, $u \in W_\infty$. So, $\bigcup_{g\in G} Wg \subseteq W_\infty$. $\square$

Two examples of this theorem follow.

Example 1. Chi [2] computed the limit set for

$$W = \text{convex}\{(1,0,0,0), (0,1,0,0), (0,0,1,0), (0,0,0,1)\}$$

and

$$p = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

using the 1-norm.

To use our theorem for this computation, we factor

$$p = PDP^{-1}$$

where $D = \text{diag}\left(1, 1, -1, \frac{1}{2}\right)$ from which we get the limit points

$$g_1 = P\ \text{diag}(1,1,-1,0)P^{-1} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$g_2 = P\ \text{diag}(1,1,1,0)P^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

From this we have

$$\begin{aligned} W_\infty &= Wg_1 \cup Wg_2 \\ &= \text{convex}\{(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 0, 1)\}. \end{aligned}$$

Example 2. Let $X_0 = \text{convex}\{(0, 0), (1, 0), (0, 1)\}$ and consider

$$X_{k+1} = X_k A + (1, 1),$$

where

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

We study the behavior of $X_0, X_1, \ldots$

Using homogeneous coordinates we can write this as

$$W_{k+1} = W_k p,$$

where

$$p = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

and

$$W_k = \{(x, 1) \colon x \in X_k\}.$$

Note that there is a nonsingular matrix R such that $P = R\,\text{diag}(1, 1, -1)R^{-1}$ so we use the norm $\|x\| = \|xR\|_\infty$ to get $\|p\| \leqslant 1$.

Now

$$p = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \quad p^2 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 2 & 0 & 1 \end{bmatrix},$$

$$p^3 = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 1 & -1 & 1 \end{bmatrix}, \quad p^4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

So $G = \{p, p^2, p^3, p^4\}$. Thus,

$$\begin{aligned} W_\infty &= Wp \cup Wp^2 \cup Wp^3 \cup Wp^4 \\ &= \text{convex}\{(1, 0, 1), (1, 1, 1), (2, 1, 1)\} \\ &\quad \cup \text{convex}\{(1, 0, 1), (2, 0, 1), (2, -1, 1)\} \\ &\quad \cup \text{convex}\{(1, 0, 1), (1, -1, 1), (0, -1, 1)\} \\ &\quad \cup \text{convex}\{(0, 0, 1), (1, 0, 1), (0, 1, 1)\} \end{aligned}$$

as shown in Fig. 1.

Note that π permutes the triangles in Fig. 1. And, dropping the third coordinate in the vectors in W_∞ (or projecting the graph into the *xy*-plane) we see the behavior of $X_0, X_1, \ldots$

We now show some further details of how the sequence $W, Wp, Wp^2, \ldots$ tends to W_∞.

Theorem 2. *$Wp^k \to W_\infty$ in the sense that if $\mathscr{O}$ is an open set containing W_∞ then there is a K such that if $k \geqslant K$ then $Wp^k \subseteq \mathscr{O}$.*

Proof. The proof is by contradiction. Suppose there is an open set $\mathscr{O}$ about W_∞, such that $Wp^j \nsubseteq \mathscr{O}$ for $j = i_1, i_2, \ldots$. Choose $w_{i_1}p^{i_1} \in Wp^{i_1}, w_{i_2}p^{i_2} \in Wp^{i_2}, \ldots$ so that $w_{i_1}p^{i_1} \notin \mathscr{O}, w_{i_2}p^{i_2} \notin \mathscr{O}, \ldots$. Let $w_{j_1}, w_{j_2}, \ldots$ be a subsequence of $w_{i_1}, w_{i_2}, \ldots$ which converges to say w and $p^{r_1}, p^{r_2}, \ldots$ a subsequence of $p^{j_1}, p^{j_2}, \ldots$ that converges to say $\bar{g} \in G$. Then $w_{r_1}p^{r_1}, w_{r_2}p^{r_2}, \ldots$ converges to $w\bar{g}$. But $w\bar{g} \in W_\infty$ which is a contradiction since $w\bar{g} \notin \mathscr{O}$. Thus, the result follows. □

The following two corollaries use the Hausdorff metric Δ and the following dilation result given in [1]: Let $\epsilon > 0$. If $A \subseteq B + \epsilon$ and $B \subseteq A + \epsilon$ then $\Delta(A, B) \leqslant \epsilon$. (Here $B + \epsilon = \{s \in S\colon\ d(s, b) \leqslant \epsilon$ for some $b \in B\}$.)

Corollary 1. *If $G = \{e\}$ then*

$$\Delta(Wp^k, W_\infty) \to 0 \quad \textit{as } k \to \infty.$$

Proof. The proof follows by showing that for any $\epsilon > 0$ there is a K such that

$$d(wp^k, we) < \epsilon \quad \text{for all } w \in W \text{ and } k \geqslant K.$$

The proof of this result is argued by contradiction as in the proof of the preceding theorem. □

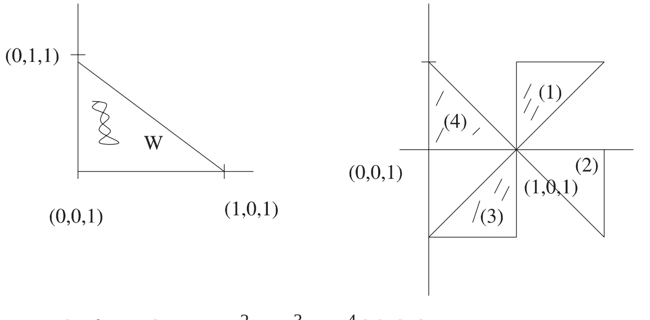

Fig. 1. Graph of W and Wp, Wp^2, Wp^3, Wp^4 labeled (1), (2), (3), (4), respectively.

A result that appears in [2,3] follows.

Corollary 2. *If* $Wp \subseteq W$ *then* $\Delta(Wp^k, W_\infty) \to 0$ *as* $k \to \infty$.

Proof. It is clear that $Wp^{k+1} \subseteq Wp^k$ for all k. It follows that $W_\infty \subseteq Wp^k$ for all k. And, by Theorem 3 we know that given any $\epsilon > 0$ there is a k such that $Wp^k \subseteq W_\infty + \epsilon$. Thus, the corollary follows. □

A theorem giving a more exacting description of the behavior of the sequence $Wp, Wp^2, \ldots$ utilizes the following lemma.

Lemma 3. *Given* $\epsilon > 0$ *and a* $g \in G$ *there is a* $k > 0$ *such that*

$$d(wp^k, wg) < \epsilon \quad \textit{for all } w \in W.$$

Proof. The proof is by contradiction. Thus, suppose there is an $\epsilon > 0$ and a $g \in G$ so that for each k there is a $\bar{w} \in W$ so that

$$d(\bar{w}p^k, \bar{w}g) \geqslant \epsilon.$$

Now let $p^{i_1}, p^{i_2}, \ldots$ be a convergent sequence which converges to g. Now for each i_k, let $w_{i_k} \in W$ be such that

$$d(w_{i_k}p^{i_k}, w_{i_k}g) \geqslant \epsilon.$$

Let $w_{j_1}, w_{j_2}, \ldots$ be a subsequence of $w_{i_1}, w_{i_2}, \ldots$ which converges to say $\bar{w}$. Since d is continuous, it follows that

$$d(\bar{w}g, \bar{w}g) \geqslant \epsilon$$

a contradiction from which the result follows. □

Theorem 3. *Suppose* $\mathscr{T}_p \leqslant 1$. *Given* $\epsilon > 0$ *and* $g \in G$ *there is a K so that if* $k \geqslant 0$ *then*

$$\Delta(Wp^{K+k}, Wgp^k) < \epsilon.$$

Proof. By Lemma 3 we know that there is a K such that

$$d(wp^K, wg) < \epsilon \quad \text{for all } w \in W.$$

Thus, since $\mathscr{T}_p \leqslant 1$,

$$d(wp^{K+k}, wgp^k) < \epsilon \quad \text{for all } w \in W.$$

This is sufficient to show that

$$\Delta(Wp^{K+k}, Wgp^k) < \epsilon. \qquad \square$$

This theorem says that for large k, Wp^k is close, in the Hausdorff metric, to some Wg, the g perhaps changing with k. (Note, by Lemma 1, $gp^k \in G$.)

The following example can be helpful.

Example 3. Let

$$p=\left\{\begin{bmatrix}0&1\\1&0\end{bmatrix},\frac{1}{2}\begin{bmatrix}1&0\\0&1\end{bmatrix}\right\}.$$

Using the 1-norm, as in (v), $\|p\|\leqslant 1$.

Consider the difference equation

$$S_{k+1}=S_k p,$$

where $W=\{(1,0)\}$. A few calculations,

$$\begin{aligned}
p&=\left\{\begin{bmatrix}0&1\\1&0\end{bmatrix},\frac{1}{2}\begin{bmatrix}1&0\\0&1\end{bmatrix}\right\},\\
p^2&=\left\{\begin{bmatrix}1&0\\0&1\end{bmatrix},\frac{1}{2}\begin{bmatrix}0&1\\1&0\end{bmatrix},\frac{1}{4}\begin{bmatrix}1&0\\0&1\end{bmatrix}\right\},\\
p^3&=\left\{\begin{bmatrix}0&1\\1&0\end{bmatrix},\frac{1}{2}\begin{bmatrix}1&0\\0&1\end{bmatrix},\frac{1}{4}\begin{bmatrix}0&1\\1&0\end{bmatrix},\frac{1}{8}\begin{bmatrix}1&0\\0&1\end{bmatrix}\right\},\\
p^4&=\left\{\begin{bmatrix}1&0\\0&1\end{bmatrix},\frac{1}{2}\begin{bmatrix}0&1\\1&0\end{bmatrix},\frac{1}{4}\begin{bmatrix}1&0\\0&1\end{bmatrix},\frac{1}{8}\begin{bmatrix}0&1\\1&0\end{bmatrix},\frac{1}{16}\begin{bmatrix}1&0\\0&1\end{bmatrix}\right\},\\
&\cdots
\end{aligned}$$

shows that $G=\{p_0,p_1\}$ where

$$\begin{aligned}
p_0&=\left\{\begin{bmatrix}0&1\\1&0\end{bmatrix},\frac{1}{2}\begin{bmatrix}1&0\\0&1\end{bmatrix},\frac{1}{4}\begin{bmatrix}0&1\\1&0\end{bmatrix},\frac{1}{8}\begin{bmatrix}1&0\\0&1\end{bmatrix},\ldots,0\right\},\\
p_1&=\left\{\begin{bmatrix}1&0\\0&1\end{bmatrix},\frac{1}{2}\begin{bmatrix}0&1\\1&0\end{bmatrix},\frac{1}{4}\begin{bmatrix}1&0\\0&1\end{bmatrix},\frac{1}{8}\begin{bmatrix}0&1\\1&0\end{bmatrix},\ldots,0\right\}.
\end{aligned}$$

(0 is added to assure p_0 and p_1 are compact.)

Thus,

$$\begin{aligned}
W_\infty&=Wp_0\cup Wp_1\\
&=\left\{(0,1),\left(\frac{1}{2},0\right),\left(0,\frac{1}{4}\right),\left(\frac{1}{8},0\right),\ldots\right\}\\
&\quad\cup\left\{(1,0),\left(0,\frac{1}{2}\right),\left(\frac{1}{4},0\right),\left(0,\frac{1}{8}\right),\ldots\right\}.
\end{aligned}$$

The first few iterates of this set are shown in Fig. 2 where squares are used for Wp_0 and circles for Wp_1. In viewing $W, Wp, \ldots$ we see that odd iterates tend to Wp_0 while even iterates tend to Wp_1. Note also that if we replace $\frac{1}{2}$ by w, $0\leqslant w<1$, in the definition of p, then W_∞ is infinite except at $w=0$.

An example requiring no computation follows.

Example 4. Let P be an $n \times n$ positive (entrywise) matrix and

$$p = \{A\colon\ A \text{ is an } n \times n \text{ doubly stochastic matrix and } A \geqslant P\,(\text{entrywise})\}.$$

Using the techniques in [4] it can be shown that $p,\ p^2, \ldots$ converges to $\{J\}$ where J is an $n \times n$ matrix with $\frac{1}{n}$ as each entry. Thus, if W is any compact set of stochastic vectors then

$$Wp, Wp^2, \ldots$$

converges to $W\{J\}$.

The following corollary, in special situations, appears in [2–4] as well as in other writings involving difference equations.

Corollary 3. $W_\infty p = \bigcup_{g \in G} W(gp) = W_\infty$.

Proof. Follows from Theorem 1 and Lemma 1. $\square$

Corollary 3 indicates that although π may contract W, it does not contract W_∞. So, for example, if W_∞ is a fractal, π will not contract it. More specifically we have the following result about W_∞ which we feel is new.

Theorem 4. *If* $\mathscr{T}_p \leqslant 1$ *then* π *is an isometry on* W_∞.

Proof. Let $\bar{x}, \bar{y} \in S$. Set $x = \bar{x}g,\ y = \bar{y}g$ where $g \in G$. Let $p^{i_1}, p^{i_2}, \ldots$ be a sequence that converges to g.

Note that

$$0 \leqslant d(\bar{x}p^{i_k}, \bar{y}p^{i_k}) - d(\bar{x}p^{i_k}p, \bar{y}p^{i_k}p) \leqslant d(\bar{x}p^{i_k}, \bar{y}p^{i_k}) - d(\bar{x}p^{i_{k+1}}, \bar{y}p^{i_{k+1}}).$$

Taking the limit as $k \to \infty$ yields

$$0 \leqslant d(\bar{x}g, \bar{y}g) - d(\bar{x}gp, \bar{y}gp) \leqslant d(\bar{x}g, \bar{y}g) - d(\bar{x}g, \bar{y}g),$$

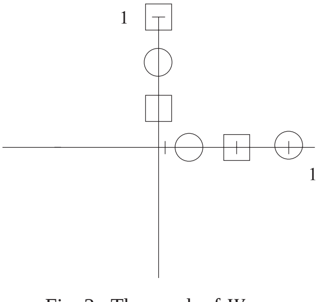

Fig. 2. The graph of W_∞.

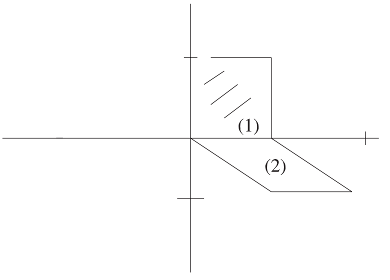

Fig. 3. Graph of W_∞. Here Wp, Wp^2 are labeled (1) and (2), respectively.

so

$$d(\bar{x}gp, \bar{y}gp) = d(\bar{x}g, \bar{y}g).$$

Now suppose that $x = \bar{x}g_1$, $y = \bar{y}g_2 \in S_\infty$ where $g_1, g_2 \in G$. Then, using the special case above,

$$d(\bar{x}g_1 p, (\bar{y}g_2 g_1^{-1})g_1 p) = d(\bar{x}g_1, (\bar{y}g_2 g_1^{-1})g_1).$$

So

$$d(\bar{x}g_1 p, \bar{y}g_2 p) = d(\bar{x}g_1, \bar{y}g_2).$$

The result follows. $\square$

It is interesting to note that the isometry depends on the metric d used for $\mathscr{T}_p$.

Example 5. Let

$$p = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}.$$

Since the eigenvalues of p are 1 and -1, as in Example 2, there is a norm $\|\cdot\|$ so that $\|p\| \leqslant 1$.

Let $W = \text{convex}\{(0,0), (1,0), (0,1), (1,1)\}$, a square. Since $p^2 = I$ it follows that

$$W_\infty = Wp \cup Wp^2$$

as shown in Fig. 3.

Note that $(Wp)p = Wp^2$ while $(Wp^2)p = Wp$ thus multiplying by p interchanges the parallelograms labeled (1) and (2). Thus, it is clear that π does not preserve distances between points in W, using the 2-norm.

References

[1] M. Barnsley, Fractals Everywhere, Academic Press, New York, 1988.
[2] M.-H. Chi, On the long run behavior of Markov chains, Linear Algebra Appl. 244 (1996) 111–121.
[3] D.J. Hartfiel, On the limit of stochastic products, Proceedings of the American Mathematical Society 81 (1981) 201–206.

[4] D.J. Hartfiel, Markov Set-Chains, LNM 1695, Springer, Berlin, 1998.
[5] A.B. Paalman-DeMiranda, Topological Semigroups, Mathematisch Centrum, Amsterdam, 1964.
[6] J.T. Sandefur, Discrete Dynamical Systems, Clarendon Press, Oxford, 1990.
[7] E. Seneta, Nonnegative Matrices and Markov Chains, second ed., Springer, Berlin, 1981.